International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V70I4P103

Volume 70 Issue 4, 19-24, April 2022

Original Article

Explainable AI – The Errors, Insights, and Lessons of AI

Vihaan Luthra[1]

[1]*School of Management, University of Colombo, Colombo, Arizona, USA.*

Received: 07 March 2022
Revised: 23 April 2022
Accepted: 25 April 2022

Abstract *- A longer-term challenge for maintaining AI's benefits is understanding how the technology can be used to create value for people and society. As AI systems perform more complex tasks, they will also become better at optimizing their performance, which could lead to undesirable outcomes if not properly managed. For example, an AI system deployed in a financial market could learn how to manipulate prices. To ensure that artificial intelligence technologies continue to benefit humanity, we need to focus on three key areas: research into making these systems reliable and beneficial, preserving our values, and managing the risks associated with these technologies. It is also important to focus on preserving our values as AI technologies advance. As these systems get better at performing tasks, they could begin to diverge from our values. Finally, because of these technologies' potential to become more powerful as they increase in capability, we need to prepare for dangerous outcomes. As AI capabilities advance, they will become better at optimizing their performance, even if this means acting in a way that diverges from human preferences or values. This could cause problems if there are significant differences between what machines optimize and humans value. To ensure that advanced AI systems continue to serve us well into the future, we need to actively study ways of controlling them so that they share our goals and do not have unexpected effects.*

Keywords *- Artificial Intelligence, Neural networks, Natural Language Processing, Virtual Assistant.*

I. INTRODUCTION

This paper aims to provide some perspective on the potential impact of advanced AI systems and how they could affect different areas of society. To do so, we first give some background on progress in AI research over the past several years. Next, we look at how specific capabilities are being applied today, focusing on three areas where there have been promising developments: virtual assistants for knowledge work [1], self-driving cars, and improved recommender systems [2]. We then examine future challenges resulting from more advanced AI capabilities, such as deploying large-scale machine learning systems across industry sectors and automating complex tasks such as medical diagnosis or legal advice. Finally, we consider how these technologies can contribute to creating value for people and society.

Progress in AI research over the past several years has been remarkable, with significant advances being made in areas such as machine learning [4], natural language processing [5], and computer vision [6]. This progress is now starting to be applied practically across several different domains. This section focuses on three specific areas where there have been promising developments: virtual assistants [7] for knowledge work, self-driving cars, and improved recommender systems.

One area where AI is making an impact is in the field of virtual assistants. Virtual assistants are programs that help people with tasks such as scheduling appointments, researching topics, or sending emails. They use natural language processing (NLP) to understand the user's requirements and then act based on that understanding.

II. MOTIVATION

One of the earliest virtual assistants was Apple's Siri, introduced in 2011. Since then, many other companies have developed their virtual assistant programs, including Google Now, Microsoft Cortana, and Amazon Echo. Millions of people around the world are now using virtual assistants. Virtual assistants are starting to be used in several domains beyond personal tasks. They are being used in knowledge work tasks such as research and data entry. A recent study by Salesforce [8] found that over 60% of knowledge workers currently use or plan to use a virtual assistant in the next year. This is because virtual assistants can be programmed to complete research, document summarization, and data entry tasks. For example, virtual assistants can read through documents online or listen to voice recordings of interviews with customers. They are also becoming better at understanding the user's requirements, enabling them to provide personal recommendations about what information is relevant to the user.

Another area where AI has made significant progress in self-driving cars. Researchers demonstrated that self-driving cars were possible back in 2005 when a team from CMU [9] won the DARPA Grand Challenge by having their autonomous vehicle complete a 150-mile course through desert scrub. Since then, there has been rapid progress in this field driven by advances in machine learning techniques for object recognition and navigation.

A third area where AI impacts is recommender systems (also called personalized ranking [10]). Recommender systems are used by companies such as Netflix, Amazon, and Spotify to recommend products (e.g., TV shows, books, and music) to their users. Recommender systems make recommendations based on a user's past behavior [11] and the behavior of other similar users; they attempt to find items that the user will like but would not have found otherwise.

III. REPORTED WORK

According to all the literature polls, most of them use BC to maintain security and privacy, while some use AI approaches. Only a few studies that we are aware of have considered AI's use with BC to maintain privacy. However, there is no comprehensive description of the processes that can be used to achieve privacy. The planned poll primarily covers the last four years (2015 to date). BC uses and problems were outlined by Madduri [3] and Ayyagari [12]. SC was introduced, and its uses were discussed [13]. Each successful machine learning trained model [14] – [16] proposed an incentive mechanism and focused on biomedical research using AI and BC. h for BC. The integration of AI and BC focused on and developed a k-anonymity-based safe authentication approach [17] – [23].

IV. PHASES OF AI

The impact of artificial intelligence on society is a widely debated topic. Some people believe that AI will have a positive impact, while others believe it will have a negative impact. There are many areas where AI has an impact, so it isn't easy to make a general statement about its overall effect. There are some areas where the potential impacts are more clear-cut.

One area where there is concern about AI's impact is employment. We are already seeing this due to increased workforce automation over the past few years. Advances in AI could lead to much greater automation that requires fewer - or possibly no - human workers. This has led many experts to predict a jobless future, where there aren't enough jobs for everyone who wants them. However, not all experts agree with this pessimistic view. Some people believe that while AI may displace some jobs currently done by humans, it will also create new types of jobs that don't exist today. They believe that technology always creates more jobs than it destroys, and because artificial intelligence is very different from previous technological advancements, it is likely to follow the same pattern.

Another area where there is concern about the impact of AI is in the field of ethics. With the rapid advancement of AI, it is becoming increasingly difficult to determine what is and isn't ethical. For example, should a self-driving car be programmed to save the lives of its passengers, even if that means sacrificing the lives of pedestrians? What about a surgical robot that has been permitted to operate on patients? Who is responsible - the robot or the human who programmed it if there is a mistake? These are just some difficult questions that we will need to answer as AI becomes more sophisticated.

This latter category includes the question of how AI will impact society that was asked in this post. The naive, binary framing made this an easy target for criticism: However, rather than just trying to find fault with this or any other article, let's make a serious attempt at addressing the underlying topic. This is one of many questions related to the future impact of artificial intelligence on our world. Indeed, I think it might be among the most important, given how essential human labor is to consumer economies worldwide. And while there are disagreements about what will happen in 10-20 years when AI begins to have a significant economic impact, I don't think anyone has anything close to a confident answer about what will happen 25-50 years from now.

It's important to remember that it took humanity centuries to figure out the best way to organize itself and our economy in a consistent way with our nature and allows for the best. Even if we had a perfect understanding of the effects of AI on society, it would not be very intelligent to think we could make an accurate prediction about what the world will look like even 25 years from now. So rather than trying futilely to answer this question, let's focus on some of the issues that arise when we ask it.

One concern is that automation powered by AI will replace jobs currently done by human beings. This has been happening for a few decades now, as machines have been taking over an increasing number of roles that used to be done by humans. However, this process has largely been confined to the manufacturing industry and a handful of service sectors. Most jobs still require some form of talent or education that machines haven't yet mastered.

There is significant disagreement in terms of what types of jobs are most likely to be affected in the future. It depends on how you define "automation" versus "artificial intelligence" - they are not necessarily the same thing, but they do overlap significantly. It seems logical for low-skill work like assembly line manufacturing to be automated eventually. At the same time, high-level thinking will remain beyond the capability of machines for many years to come. Some economists argue that it is precisely low-level human jobs that are most likely to be replaced by AI soon: In contrast, I've read claims of influential people like Stephen Hawking and Bill Gates that seem to indicate a

belief that artificial intelligence will replace many types of high-skill work as well:

This may sound surprising since today's AI seems better at solving problems than humans, rather than taking over entire professions. There is no robot with the ability to do every task involved with being a doctor or lawyer. But with technology moving so quickly, we can't rule out the possibility that the machine learning algorithms used for such narrow tasks today could be repurposed for general-purpose problem-solving. It's hard to know when, but it's possible that in a few decades, the typical professional will be working in tandem with a machine that can do a lot of their job for them - or perhaps even take over completely.

The most alarming prospect is what this would mean for people who make their living providing skilled labor: doctors, lawyers, engineers, marketers, etc. These jobs require years of training and countless hours of experience to master - what happens if they become obsolete overnight? The most likely outcome is that humans will retain control of non-routine problems while machines handle what is predictable and repetitive. But evidence from the past suggests there may be more dramatic changes ahead.

There were no machines to take over tasks that required manual labor for most of human history. Then the Industrial Revolution happened, and humans started building machines that let them work faster than ever before, taking away jobs from many lower-skilled workers and freeing up other types of jobs for those who either had the necessary technical skills or could afford to hire various forms of skilled labor:

This pattern repeated itself with every technological development throughout modern history - as technology became more advanced, it often made certain professions obsolete while simultaneously opening new possibilities. The same has been true even as AI becomes more sophisticated: as we've built better tools powered by machine learning algorithms, people have used those tools to produce things that didn't exist before, like self-driving cars and AlphaGo. It's hard to say what the future of work will look like. But looking at the evidence from history, it seems safe to say that there will be more change ahead and that many jobs currently considered skilled labor may eventually be replaced by machines. It's also likely that humans will continue to find new ways to use technology to create value. We don't know what those new jobs will be yet.

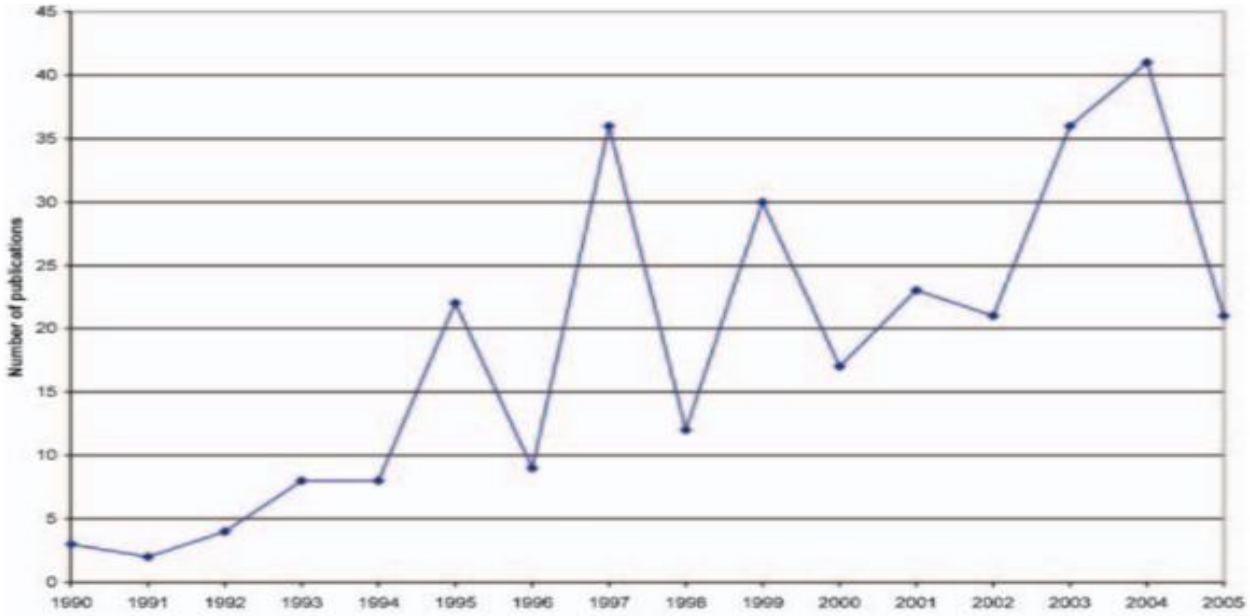

Fig. 1 A Parody of Neurobiology

Inputs can be delivered at the top in Fig. 1, and external outputs can be taken from some amplifiers. This completes our artificial neural network model. It is a parody of neurobiology but already encompasses many features that differentiate digital-logical circuits from neurobiological ones.

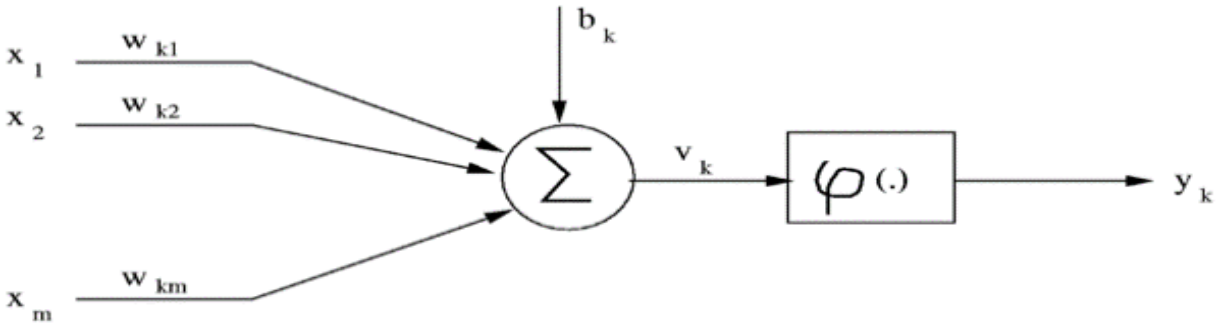

Fig. 2 Another Possible Arrangement

Fig. 2 identifies some of the amplifiers. This completes our artificial neural network model. It is a parody of neurobiology but already encompasses many features that differentiate digital-logical circuits from neurobiological ones. It can be taken from some of the amplifiers. Fig 3 completes our artificial neural network model. It is a parody of neurophysiology but already encompasses many features that differentiate digital-logical circuits from neurobiological ones.

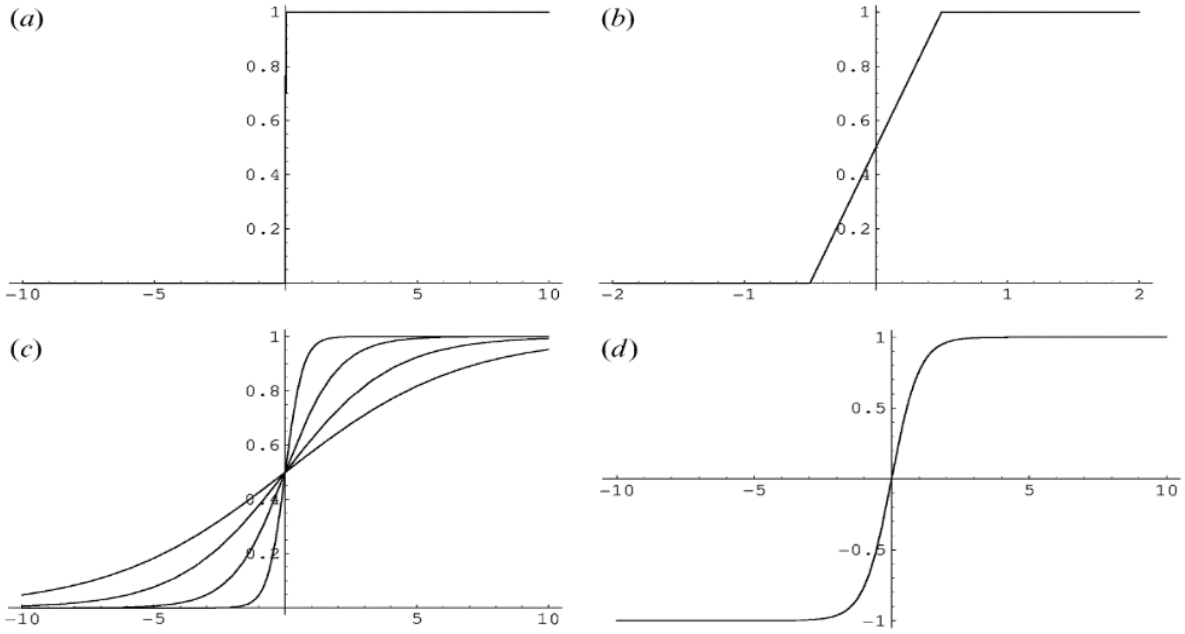

Fig. 3 Complete Circuit Diagrams for the Four-Layer Paradigm

V. EXPERIMENTAL RESULTS

In this section, the authors have presented experimental results of the artificial neural network model that encompasses many features that differentiate digital-logical circuits from neurobiological ones. Spectral normalization is also applied to the generator. Initiating a self-attention and generating premier image quality has been comprehended in the image generation of the specified class. The network using self-attention has local information and a global view. Progress inefficient learning methods and transfer learning are also important for supervised learning of diagnostic medical imaging. Before supervised learning using existing images for AI, training with general image data sets is important. This can be too the case for the ART1 analysis. Although ART1, with half the first number of features, did nearly as well or somewhat way better, we chose to show that the leftover portion of the investigation utilizes the original feature's instep. One reason is that the central components have by and large small clinical interpretability. Given the constrained information, it would have been unreasonable to attempt to go beyond the straight reliance. Without solid proof of linear reliance between the highlights, we contend that obtaining ART1 components that can capture the larger part of the information fluctuation comes at the chance of reduced sensitivity and may not fundamentally interpret correct labeling. The beat three classifiers remained the same with both ART1 and PCA highlights as in Fig. 4 and Fig. 5. In the future, parameters such as blood and hereditary tests may include more visionary capacity. With this more prominent amount of information, the choice of strategy highlights determination and classification.

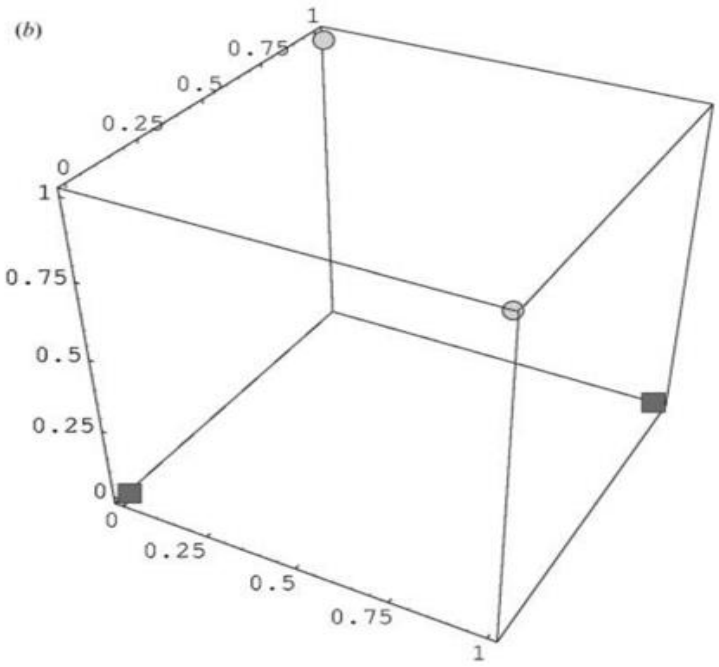

Fig. 4 Complex Arrangement with Several Layers

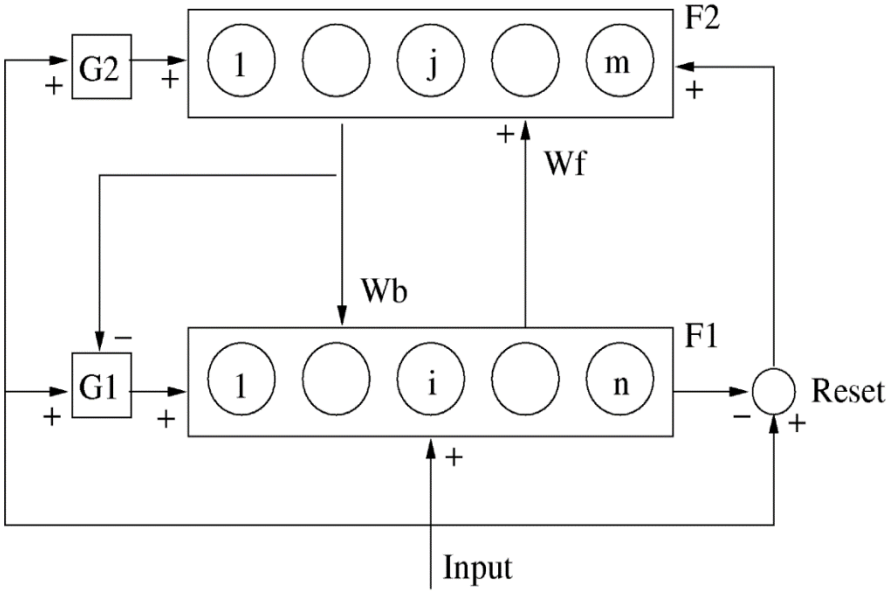

Fig. 5 ART1 Model

VI. CONCLUSION

In conclusion, there has been a lot of progress made in AI over the past few years, and this progress is having a significant impact on society. In the next article in this series, we will look at how AI is being used in healthcare.

REFERENCES

[1] Kastaniotis I, Theodorakopoulos C, Theoharatos GE and Fotopoulos S., A framework for gait-based recognition using Kinect. Pattern Recognition Letters, 68 (2015)327–335.

[2] Ayyagari, M. R., A framework for analytical CRM assessments challenges and recommendations. International Journal of Information, Business, and Management, 13(2) (2021) 108-121.

[3] Madduri, A. Human Gait Recognition using Discrete Wavelet and Discrete Cosine and Transformation Based. International Journal of Computer Trends and Technology. (IJCTT), 69 (6) 22-27

[4] Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G., A survey of deep learning and its applications: a new paradigm to machine learning. Archives of Computational Methods in Engineering, (2019) 1-22.

[5] Kumar, G., Thakur, K., & Ayyagari, M. R., MLEsIDSs: machine learning-based ensembles for intrusion detection systems—a review. The Journal of Supercomputing, (2020) 1-34.

[6] Madduri, A. Content-based Image Retrieval System using Local Feature Extraction Techniques. International Journal of Computer Applications, 975 8887.

[7] Ayyagari, R. M., & Atoum, I., CMMI-DEV Implementation Simplified. International Journal of Advanced Computer Science and Applications, 10(4) (2019) 445-459.

[8] Mehraj, H., Jayadevappa, D., Haleem, S. L. A., Parveen, R., Madduri, A., Ayyagari, M. R., & Dhabliya, D., Protection motivation theory uses multi-factor authentication to provide security over social networking sites. Pattern Recognition Letters, 152 (2021) 218-224.

[9] Ayyagari, M. R. Classification of Imbalanced Datasets using One-Class SVM, k-Nearest Neighbors, and CART Algorithm.
[10] Ayyagari, M. R., Kesswani, N., Kumar, M., & Kumar, K. ., Intrusion detection techniques in a network environment: a systematic review. Wireless Networks, (2021) 1-17.
[11] Atoum, I., & Ayyagari, M. R., Effective Semantic Text Similarity Metric Using Normalized Root Mean Scaled Square Error. Journal of Theoretical and Applied Information Technology, 97 (2019) 3436-3447.
[12] Ayyagari, M. R. Efficient Driving Forces to CMMI Development using Dynamic Capabilities. International Journal of Computer Applications, 975 8887.
[13] Ayyagari, M. R., Integrating Association Rules with Decision Trees in Object-Relational Databases. arXiv preprint arXiv:1904.09654. (2019).
[14] Ayyagari, M. R., iScrum: Effective Innovation Steering using Scrum Methodology. Int. J. Comput. Appl, 178(10) (2019) 8-13.
[15] Ayyagari, M. R. Cache Contention on Multicore Systems: An Ontology-based Approach.(2019). arXiv preprint arXiv:1906.00834.
[16] Ayyagari, M. R., & Atoum, I., Understanding Customer Voice of Project Portfolio Management Software. Int. J. Adv. Comput. Sci. Appl, 10(5) (2019) 51-56.
[17] Ayyagari, M. R., Integrating Association Rules with Decision Trees in Object-Relational Databases. arXiv preprint arXiv:1904.09654. (2019).
[18] Ye and Wen YE, Gait recognition based on DWT and SVM. Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, 3 (2007) 1382–1387.
[19] Gupta D and Choubey S., Discrete wavelet transform for image processing. International Journal of Emerging Technology and Advanced Engineering, 4(3) (2015) 598-602.
[20] Fan Z, Jiang J, Weng S, He Z, and Liu Z., Human gait recognition based on discrete cosine transform and linear discriminant analysis. Proceedings of the International Conference on Signal Processing, Communications and Computing (IC- SPCC), (2016) 1–6.
[21] Atoum, Issa, and Ayyagari, Maruthi Rohit., Effective Semantic Text Similarity Metric Using Normalized Root Mean Scaled Square Error. Journal of Theoretical and Applied Information Technology. 97 (2019)3436-3447.
[22] Pal M and Mather PM ., An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment, 86(4) (2003) 554– 565.
[23] Albert J, Aliu E, Anderhub H, Antoranz P, Armada A, Asensio M, Baixeras C, Barrio J, Bartko H, Bastieri D ., Implementation of the random forest method for the imaging atmospheric Cherenkov telescope magic. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 588(3) (2008) 424-432.
[24] Belgium M and Dr˘agut L (2016) Random Forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114 (2016) 24–31.
[25] Chattopadhyay P, Sural S, and Mukherjee J ., Frontal gait recognition from incomplete sequences using an RGB-D camera. IEEE Transactions on Information Forensics and Security, 9(11) (2014a) 1843–1856.
[26] Ayush Jain, Prathamesh Patil, Ganesh Masud, Prof. Sunitha Krishnan, Prof. Vijaya Bharathi Jagan. Detection of Sarcasm through Tone Analysis on video and Audio files: A Comparative Study On Ai Models Performance. SSRG International Journal of Computer Science and Engineering 8(12) (2021) 1-5.
[27] S.Kalpakha, G.Kalaiselvan, T.Aravindh Krishna., Innovative Digital Customer Engagement and Experience in Car Retail using Augmented and Virtual Reality. SSRG International Journal of Computer Science and Engineering 5(12) (2018) 18-23.